

车辆的环境检测方法、装置、设备、车辆、介质及产品

交叉引用

本申请要求于 2025 年 1 月 13 日提交中国专利局、申请号为 202510047555.X、发明名称为“车辆的环境检测方法、装置、设备、车辆、介质及产品”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及但不限于汽车技术领域，例如涉及一种车辆的环境检测方法、装置、设备、车辆、介质及产品。

背景技术

在对车辆的周围环境进行检测时，充分利用车载各视角图像信息来提升对目标和环境的综合感知定位能力。在行、泊车功能需求的前提支持下，车辆均会配备足以覆盖车周 360° 视角的周视相机与环视相机，实现对车周例如中近距离的探测覆盖。

公开内容

以下是对本文详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

本公开实施例提供一种车辆的环境检测方法、装置、设备、车辆、介质及产品。

本公开实施例采用的技术方案如下：

一种车辆的环境检测方法，包括：分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征；基于检测任务对应的任务类型，确定第一查询信息；基于周视图像特征和环视图像特征，确定第一键信息和第一值信息；利用注意力机制，对第一查询信息、第一键信息和第一值信息进行处理，得到目标融合特征；及，基于目标融合特征，对车辆的周围环境进行检测，确定与检测任务对应的检测结果。

根据上述技术手段，分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征；之后根据检测任务对应的任务类型，确定第一查询信息，以及根据周视图像特征和环视图像特征，确定检测任务对应的第一键信息和第一值信息；利用注意力机制，对第一查询信息、第一键信息和第一值信息进行处理，得到目标融合特征；之后，根据目标融合特征，对车辆的周围环境进行检测，确定与检测任务对应的检测结果。这样，一方面，利用注意力机制以使周视图像特征和环视图像特征一体化输入输出，减少了独立运行带来的资源抢占与延时等待，有效提升信息处理效率，并且，通过以周视图像特征和环视图像特征构建注意力机

制中的键信息和值信息，可以将周视图像特征、环视图像特征更好地融合利用，从而可以提升环境检测过程中的感知能力，以提高环境检测的准确率；另一方面，利用注意力机制，对任务相关信息、周视图像特征、环视图像特征进行稀疏化表达，从而减少了特征融合过程中的计算量和内存占用。

可选地，周视图像特征包括二维周视特征和三维周视位置特征，环视图像特征包括二维环视特征和三维环视位置特征；分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征，包括：对周视图像和环视图像进行二维编码，确定二维周视特征和二维环视特征；及，对周视图像和环视图像进行三维编码，确定三维周视位置特征和三维环视位置特征。

根据上述技术手段，周视图像特征包括二维周视特征和三维周视位置特征，环视图像特征包括二维环视特征和三维环视位置特征；对周视图像和环视图像进行二维编码，确定二维周视特征和二维环视特征；对周视图像和环视图像进行三维编码，确定三维周视位置特征和三维环视位置特征。这样，可以使周视图像特征和环视图像特征中包括完整的二维的图像信息和三维的空间位置信息，提高环境检测的准确性。

可选地，基于检测任务对应的任务类型，确定第一查询信息，包括：基于检测任务对应的任务类型和各历史时刻检测任务对应的检测结果，确定检测任务对应的第二查询信息、第二键信息和第二值信息；及，利用注意力机制，对第二查询信息、第二键信息和第二值信息进行处理，得到第一查询信息。

根据上述技术手段，通过根据检测任务对应的任务类型和各历史时刻该检测任务对应的检测结果，确定该检测任务对应的第二查询信息、第二键信息和第二值信息；利用注意力机制，对第二查询信息、第二键信息和第二值信息进行处理，得到第一查询信息，可以将当前时刻的任务相关信息和各历史时刻的任务相关信息融合。

可选地，基于检测任务对应的任务类型和各历史时刻检测任务对应的检测结果，确定检测任务对应的第二查询信息、第二键信息和第二值信息，包括：基于各历史时刻与当前时刻车辆的位姿转换关系，对各历史时刻检测任务对应的检测结果进行重构，确定各历史时刻检测任务对应的重构特征；基于检测任务对应的任务类型和第一历史时刻检测任务对应的重构特征，确定当前时刻检测任务对应的初始重构特征；其中，第一历史时刻为当前时刻的上一时刻；将当前时刻检测任务对应的初始重构特征确定为第二查询信息；及，基于各第二历史时刻检测任务对应的重构特征和当前时刻检测任务对应的初始重构特征，确定第二键信息和第二值信息；第二历史时刻为除第一历史时刻之外的历史时刻。

根据上述技术手段，基于各历史时刻与当前时刻车辆的位姿转换关系，对各历史时刻检测任务对应的检测结果进行重构，确定各历史时刻检测任务对应的重构特征，以及根据检测任务对应的任务类型和第一历史时刻检测任务对应的重构特征，确定当前时刻检测任务对应的初始重构特征；第一历史时刻为当前时刻的上一时刻；之后，将当前时刻检测任务对应的初始重构特征确定为第二查询信息；之后，基于各第二历史时刻检测任务对应的重构特征和当前时刻检测任务对应的初始重构特征，确定第二键信息和第二值信息；第二历史时刻为

除第一历史时刻之外的历史时刻。这样，第二查询信息包括当前时刻检测任务对应的初始重构特征，第二键信息和第二值信息包括各第二历史时刻检测任务对应的重构特征和当前时刻检测任务对应的初始重构特征，如此，可以在第一查询信息中融合当前时刻的任务相关信息和各历史时刻的任务相关信息，进而可以在目标融合特征中建立当前时刻的任务相关信息和各历史时刻的任务相关信息之间的关联，以提高环境检测的准确性。

可选地，基于各历史时刻与当前时刻车辆的位姿转换关系，对各历史时刻检测任务对应的检测结果进行重构，确定各历史时刻检测任务对应的重构特征，包括：基于位姿转换关系，将各历史时刻检测任务对应的检测结果对齐至当前时刻，确定各历史时刻检测任务对应的更新检测结果；及，基于位姿转换关系、各历史时刻与当前时刻之间的时间差、检测任务对应的运动速度，对各历史时刻检测任务对应的更新检测结果进行重构，确定各历史时刻检测任务对应的重构特征。

根据上述技术手段，通过根据位姿转换关系，将各历史时刻所述检测任务对应的检测结果对齐至当前时刻，确定各历史时刻检测任务对应的更新检测结果；根据位姿转换关系、各历史时刻与当前时刻之间的时间差、检测任务对应的运动速度，对各历史时刻检测任务对应的更新检测结果进行重构，确定各历史时刻检测任务对应的重构特征。这样，将历史时刻检测任务对应的检测结果对齐至当前时刻并进行重构，以使当前时刻任务相关信息和历史时刻任务相关信息在语义上保持连续性和一致性。

可选地，基于检测任务对应的任务类型和第一历史时刻检测任务对应的重构特征，确定当前时刻检测任务对应的初始重构特征，包括：基于第一历史时刻检测任务对应的重构特征，确定当前时刻检测任务对应的第一初始重构特征；基于检测任务对应的任务类型，随机生成当前时刻检测任务对应的第二初始重构特征；及，将第一初始重构特征和第二初始重构特征进行拼接，确定当前时刻对应的初始重构特征。

根据上述技术手段，根据第一历史时刻检测任务对应的重构特征，确定当前时刻检测任务对应的第一初始重构特征；之后，根据检测任务对应的任务类型，随机生成当前时刻检测任务对应的第二初始重构特征；之后，将第一初始重构特征和第二初始重构特征进行拼接，确定当前时刻对应的初始重构特征。这样，一方面，利用第一历史时刻检测任务对应的重构特征初始化可以使当前时刻任务相关信息和历史时刻任务相关信息保持一定的连续性，有助于减少因随机初始化而导致不稳定性，另一方面，利用随机初始化学习到更多的多样性信息。

可选地，所述基于检测任务对应的任务类型，随机生成当前时刻检测任务对应的第二初始重构特征，包括：基于检测任务对应的任务类型，随机生成当前时刻检测任务对应的初始特征；及，对与任务类型对应的初始特征进行重构，确定当前时刻检测任务对应的第二初始重构特征。

根据上述技术手段，根据检测任务对应的任务类型，随机生成当前时刻检测任务对应的初始特征；对检测任务对应的初始特征进行重构，确定当前时刻检测任务对应的第二初始重构特征。这样，对检测任务对应的初始特征进行重构保持与各历史时刻特征的一致性。

可选地，基于所述检测任务对应的任务类型，随机生成当前时刻检测任务对应的初始特征，包括以下至少之一：在检测任务中包括运动目标检测任务的情况下，对三维空间的坐标进行均匀采样，确定运动目标检测任务对应的初始特征；在检测任务中包括道路分割检测任务的情况下，基于设定的车辆可通行区域中各子区域的中心位置的坐标，确定道路分割检测任务对应的初始特征；及，在检测任务中包括车道线检测任务的情况下，对二维空间的坐标进行均匀采样，确定车道线检测任务对应的初始特征。

根据上述技术手段，在检测任务中包括运动目标检测任务的情况下，对三维空间的坐标进行均匀采样，确定运动目标检测任务对应的初始特征；在检测任务中包括道路分割检测任务的情况下，基于设定的车辆可通行区域中各子区域的中心位置的坐标，确定道路分割检测任务对应的初始特征；在检测任务中包括车道线检测任务的情况下，对二维空间的坐标进行均匀采样，确定车道线检测任务对应的初始特征。这样，针对不同的检测任务采用不同的初始化方式，提高了初始化的准确性。

可选地，所述基于周视图像特征和环视图像特征，确定检测任务对应的第一键信息和第一值信息，包括：将二维周视特征和二维环视特征进行拼接，确定检测任务对应的第一键信息；及，将二维周视特征、三维周视位置特征、二维环视特征和三维环视位置特征进行拼接，确定检测任务对应的第一值信息。

根据上述技术手段，将二维周视特征和所述二维环视特征进行拼接，确定检测任务对应的第一键信息；将二维周视特征、三维周视位置特征、二维环视特征和三维环视位置特征进行拼接，确定检测任务对应的第一值信息。这样，可以使得第一值信息和第一键信息中包括完整的周视图像特征和环视图像特征，提高对车辆的周围环境检测的准确性。

可选地，利用注意力机制，对第一查询信息、第一键信息和第一值信息进行处理，得到目标融合特征，包括：将第一查询信息作为注意力机制的查询矩阵、第一键信息作为注意力机制的键矩阵，生成注意力矩阵；及，基于注意力矩阵和第一值信息，确定目标融合特征。

根据上述技术手段，将第一查询信息作为注意力机制的查询矩阵、第一键信息作为注意力机制的键矩阵，生成注意力矩阵；基于注意力矩阵和第一值信息，确定目标融合特征。这样，一方面，利用注意力机制对任务相关信息、周视图像特征、环视图像特征进行稀疏化表达，从而减少了计算量和内存占用；另一方面，利用注意力机制将周视图像特征和环视图像特征进行融合，以及建立各历史时刻与当前时刻之间的关联，以提高对车辆的周围环境检测的准确性。

可选地，检测任务的数量为多个；基于检测任务对应的任务类型，确定第一查询信息，包括：针对每一检测任务，基于检测任务对应的任务类型，确定检测任务对应的查询信息；及，将各检测任务对应的查询信息进行拼接，确定第一查询信息。

根据上述技术手段，针对每一检测任务，基于检测任务对应的任务类型，确定检测任务对应的查询信息；将各检测任务对应的查询信息进行拼接，确定第一查询信息。这样，多任务融合实现对道路环境的全面感知，提高对车辆周围环境检测的准确性和鲁棒性。

可选地，该方法还包括：对车辆的雷达点云数据进行特征提取，得到点云特征；基于周视图像特征和环视图像特征，确定第一键信息和第一值信息，包括：基于周视图像特征、环视图像特征和点云特征，确定第一键信息和第一值信息。

根据上述技术手段，通过对车辆的雷达点云数据进行特征提取，得到点云特征，并根据周视图像特征、环视图像特征和点云特征，确定第一键信息和第一值信息，以使第一键信息和第一值信息中包括纯视觉的周视图像特征和环视图像特征，以及点云特征，提高对车辆的周围环境检测的准确性。

一种车辆的环境检测装置，所述环境检测装置包括：

特征提取部，配置为分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征；

第一确定部，配置为基于检测任务对应的任务类型，确定第一查询信息；

第二确定部，配置为基于周视图像特征和环视图像特征，确定检测任务对应的第一键信息和第一值信息；

处理部，配置为利用注意力机制，对第一查询信息、第一键信息和第一值信息进行处理，得到目标融合特征；及

第三确定部，配置为基于目标融合特征，对车辆的周围环境进行检测，确定与检测任务对应的检测结果。

本公开实施例提供一种计算机设备，包括存储器和处理器，存储器存储有可在处理器上运行的计算机指令，处理器执行指令时实现上述方法中的部分或全部步骤。

本公开实施例提供一种车辆，包括上述的计算机设备。

本公开实施例提供一种计算机可读存储介质，其上存储有计算机指令，该计算机指令被处理器执行时实现上述方法中的部分或全部步骤。

本公开实施例提供一种计算机程序产品，包括计算机指令，计算机指令被处理器执行时，实现上述方法中的部分或全部步骤。

在阅读并理解了附图和详细描述后，可以明白其他方面。

附图说明

图 1 为本公开实施例提出的一种车辆的环境检测方法的实现流程示意图；

图 2 为本公开实施例提出的一种利用稀疏化 HAS-Propagation Transformer 架构进行车辆环境检测的结构示意图；

图 3 为本公开实施例提出的一种利用稀疏化 HAS-Propagation Transformer 架构进行车辆环境检测的网络结构示意图；

图 4 为本公开实施例提出的一种稀疏化 HAS-Propagation Transformer 架构的 BEV 周环视动静态多任务感知方法的实现过程示意图；

图 5 为本公开实施例提出的一种相机坐标系和本车坐标系的示意图；

图 6 为本公开实施例提出的一种历史缓存信息结构示意图；

图 7 为本公开实施例提出的一种 3D 目标检测任务对应的历史缓存 Embedding 重构示意图；

图 8 为本公开实施例提出的一种道路分割检测任务对应的历史缓存 Embedding 重构示意图；

图 9 为本公开实施例提出的一种车道线检测任务对应的历史缓存 Embedding 重构示意图；

图 10 为本公开实施例提出的一种各检测任务物理对象示意图；

图 11 为本公开实施例提出的一种 HAS-Propagation Transformer 结构示意图；

图 12 为本公开实施例提出的一种车辆的环境检测装置的组成结构示意图；

图 13 为本公开实施例提出的一种计算机设备的硬件实体示意图。

具体实施方式

以下将参照附图和实施例来说明本公开实施例的实施方式，本领域技术人员可由本说明书中所揭露的内容了解本公开实施例的其他优点与功效。本公开实施例还可以通过另外不同的实施方式加以实施或应用，本说明书中的各项细节也可以基于不同观点与应用，在没有背离本公开实施例的情况下进行各种修饰或改变。实施例仅为说明本公开实施方案，而不是为了限制本公开实施例的保护范围。

以下实施例中所提供的图示仅以示意方式说明本公开实施例的基本构想，遂图式中仅显示与本公开实施例中有关的组件而非按照实际实施时的组件数目、形状及尺寸绘制，其实际实施时各组件的型态、数量及比例可为一种随意的改变，且其组件布局型态也可能更为复杂。

在实现本公开实施例的过程中，发现由于周视图像与环视图像在视野范围、畸变模式等方面存在较大差异，周视与环视针对行、泊任务一般是各自独立处理，造成车端有限的资源消耗和因抢占造成的高延时，以及对周视固有的视野盲区特征缺陷无法有效弥补。

本公开实施例提出了一种车辆的环境检测方法，如图1所示，车辆的环境检测方法包括如下步骤S101至步骤S105，其中：

步骤S101：分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征；

其中，周视图像是一种能够展现某个区域或对象周围环境的图像，这种图像可以从一个中心点出发，向

四周辐射的视觉呈现。周视图像能够捕捉到周围环境的高清细节，但是覆盖范围有限。

其中，环视图像能够捕捉并展示某个场景或对象360度全景的图像，覆盖范围大，但畸变较大。

在一些实施例中，通过周视相机，例如针孔透视相机，获取周视图像；通过环视相机，例如鱼眼相机，获取环视图像。

在一些实施例中，周视图像特征包括二维周视特征和三维环视位置特征。

在一些实施例中，环视图像特征包括二维环视特征和三维环视位置特征。

步骤S102：基于检测任务对应的任务类型，确定第一查询信息；

在一些实施例中，检测任务可以包括运动目标检测任务、道路分割检测任务和车道线检测任务。

在一些实施例中，不同的检测任务为不同的任务类型。

在一些实施例中，第一查询信息为与检测任务相关的信息。

在一些实施例中，可以根据检测任务对应的任务类型确定查询矩阵、键矩阵、值矩阵，利用混合注意力机制，确定第一查询信息。

步骤S103：基于所述周视图像特征和所述环视图像特征，确定第一键信息和第一值信息；

在一些实施例中，第一键信息包括周视二维特征和环视二维特征。

在一些实施例中，第一值信息包括周视二维特征、周视三维特征、环视二维特征和环视三维特征。

步骤S104：利用注意力机制，对所述第一查询信息、所述第一键信息和所述第一值信息进行处理，得到目标融合特征；

在一些实施例中，将第一查询信息作为查询矩阵，将第一键信息作为键矩阵，将第一值信息作为值矩阵，利用周视环视注意力机制，确定目标融合特征。

步骤S105：基于所述目标融合特征，对所述车辆的周围环境进行检测，确定与所述检测任务对应的检测结果。

在一些实施例中，将检测任务对应的目标融合特征接入对应的解码头网络，得到与检测任务对应的多个解码结果，根据解码结果对应的解码置信度进行排序，从多个解码结果中选取目标数量的目标解码结果；根据目标解码结果，确定与检测任务对应的检测结果。

在一些实施例中，将排序后的解码结果采用先进先出（FIFO）方式存入历史缓存队列，供下一时刻处理使用。

在一些实施例中，运动目标检测任务的解码可采用简单的多层感知机网络，输入维度为C，输出维度则对齐目标置信度、目标类别、中心位置、矩形框长宽高、目标朝向及目标速度等信息量，根据实际需要设定。

在一些实施例中，道路分割检测任务的解码可以分割出BEV平面上的像素级位置区域，由于每个道路查询

(Query) 对应鸟瞰图 (Bird's Eye View, BEV) 平面 $S_H \times S_W$ 个矩形块面片中的一个, 每个面片对应 $r_h \times r_w$ 像素大小, 因此可采用输入维度是C、输出维度是 $r_h \times r_w$ 加1维置信度的简单多层感知机网络; 也可采用简单的CNN网络对 $r_h \times r_w$ 进行像素分割。

在一些实施例中, 车道线检测任务采用多层感知机网络, 输入维度是C, 但其输出维度取决于对车道线的矢量化表达形式, 通常较为典型的有MapTR、PETRv2等表达方式, 甚至可采用3D车道线PersFormer方式 (如采用3D车道线), 可结合实际需要及复杂度限制进行衡量, 本公开实施例不做约束。

本公开实施例中, 分别对车辆的周视图像和环视图像进行特征提取, 得到周视图像特征和环视图像特征; 之后根据检测任务对应的任务类型, 确定第一查询信息, 以及根据周视图像特征和环视图像特征, 确定检测任务对应的第一键信息和第一值信息; 利用注意力机制, 对第一查询信息、第一键信息和第一值信息进行处理, 得到目标融合特征; 之后, 根据目标融合特征, 对车辆的周围环境进行检测, 确定与检测任务对应的检测结果。这样, 一方面, 利用注意力机制以使周视图像特征和环视图像特征一体化输入输出, 减少了独立运行带来的资源抢占与延时等待, 有效提升信息处理效率, 并且, 通过以周视图像特征和环视图像特征构建注意力机制中的键信息和值信息, 可以将周视图像特征、环视图像特征更好地融合利用, 从而可以提升环境检测过程中的感知能力, 以提高环境检测的准确率; 另一方面, 利用注意力机制, 对任务相关信息、周视图像特征、环视图像特征进行稀疏化表达, 从而减少了特征融合过程中的计算量和内存占用。

在一些实施例中, 可以基于周视图像特征和环视图像特征, 确定第一键信息和第一值信息。这样, 可以通过对纯视觉的周视图像特征和环视图像特征进行融合, 得到用于对车辆的周围环境进行检测的目标融合特征, 减少了使用其他非视觉采集设备 (例如激光雷达等) 进行检测的高昂成本, 从而降低了对车辆的周围环境进行检测的成本。

在一些实施例中, 可以基于周视图像特征、环视图像特征以及对车辆的非视觉模态感知数据 (例如雷达点云数据等) 进行特征提取得到的非视觉模态特征, 确定第一键信息和第一值信息。这样, 可以通过对视觉模态下的周视图像特征和环视图像特征、以及非视觉模态特征进行融合, 提高对环境的感知能力, 从而提高对车辆的周围环境检测的准确性。

在一些实施例中, 所述周视图像特征包括二维周视特征和三维周视位置特征, 所述环视图像特征包括二维环视特征和三维环视位置特征; 步骤S101中所述的分别对车辆的周视图像和环视图像进行特征提取, 得到周视图像特征和环视图像特征, 可以包括如下步骤S1011和步骤S1012, 其中:

步骤S1011: 对所述周视图像和所述环视图像进行二维编码, 确定所述二维周视特征和所述二维环视特征;

在一些实施例中, 分别利用两个独立的2D卷积编码主干网络对周视图像和环视图像进行特征编码, 获取二维周视特征和二维环视特征。

在一些实施例中,主干网络通常可采用但不限于有一定深度层级的如ResNet、EfficientNet、SwinTransformer、VoVNetV2等系列常用2D卷积网络结构。

在一些实施例中,所有周视图像和所有环视图像分别各自共享一套网络权重。

步骤S1012:对所述周视图像和所述环视图像进行三维编码,确定所述三维周视位置特征和所述三维环视位置特征。

在一些实施例中,根据该周视相机的畸变模型及畸变模型对应的畸变参数,结合周视相机内参,计算周视图像中各像素点在周视相机坐标系下对应的去畸变真实矢量方向;接着,在BEV空间设置合理的等间隔深度区段,并结合各像素点的矢量方向构建图像中各像素点在周视相机坐标系下不同深度区段的坐标值;然后,利用周视相机相对本车坐标系下的外参旋转和位移信息,将各像素点在不同深度区段下的坐标转换至本车坐标系下;之后,采用参考位置编码(Reference Positional Encoding, RefPE)对本车坐标系下周视图像中各像素点对应的不同深度区段坐标进行编码,得到三维周视位置特征。

在一些实施例中,根据该环视相机的畸变模型及畸变模型对应的畸变参数,结合环视相机内参,计算环视图像中各像素点在环视相机坐标系下对应的去畸变真实矢量方向;接着,在BEV空间设置合理的等间隔深度区段,并结合各像素点的矢量方向构建图像中各像素点在环视相机坐标系下不同深度区段的坐标值;然后,利用环视相机相对本车坐标系下的外参旋转和位移信息,将各像素点在不同深度区段下的坐标转换至本车坐标系下;之后,采用参考位置编码(Reference Positional Encoding, RefPE)对本车坐标系下环视图像中各像素点对应的不同深度区段坐标进行编码,得到三维环视位置特征。

本公开实施例中,周视图像特征包括二维周视特征和三维周视位置特征,环视图像特征包括二维环视特征和三维环视位置特征;对周视图像和环视图像进行二维编码,确定二维周视特征和二维环视特征;对周视图像和环视图像进行三维编码,确定三维周视位置特征和三维环视位置特征。这样,可以使周视图像特征和环视图像特征中包括完整的二维的图像信息和三维空间位置信息,提高环境检测的准确性。

在一些实施例中,步骤S102中所述的基于检测任务对应的任务类型,确定第一查询信息,可以包括如下步骤S1021和步骤S1022,其中:

步骤S1021:基于所述检测任务对应的任务类型和各历史时刻所述检测任务对应的检测结果,确定所述检测任务对应的第二查询信息、第二键信息和第二值信息;

在一些实施例中,根据检测任务对应的任务类型初始化当前时刻的初始特征,根据上一时刻检测任务对应的检测结果和当前时刻的初始特征,确定检测任务对应的第二查询信息。

在一些实施例中,根据当前时刻的初始特征和各历史时刻检测任务对应的检测结果,确定第二键信息和第二值信息;第二键信息和第二值信息一致。

步骤S1022：利用所述注意力机制，对所述第二查询信息、所述第二键信息和所述第二值信息进行处理，得到所述第一查询信息。

在一些实施例中，第二查询信息作为查询矩阵、第二键信息作为键矩阵、第二值信息作为值矩阵，输入到混合注意力机制中，输出第一查询信息。

本公开实施例中，通过根据检测任务对应的任务类型和各历史时刻所述检测任务对应的检测结果，确定检测任务对应的第二查询信息、第二键信息和第二值信息；利用注意力机制，对第二查询信息、第二键信息和第二值信息进行处理，得到第一查询信息，可以将当前时刻的任务相关信息和各历史时刻的任务相关信息融合。

在一些实施例中，步骤S1021中所述的基于所述检测任务对应的任务类型和各历史时刻所述检测任务对应的检测结果，确定所述检测任务对应的第二查询信息、第二键信息和第二值信息，可以包括如下步骤S10211至步骤S10214，其中：

步骤S10211：基于各所述历史时刻与当前时刻所述车辆的位姿转换关系，对各所述历史时刻所述检测任务对应的检测结果进行重构，确定各所述历史时刻所述检测任务对应的重构特征；

在一些实施例中，从历史缓存队列中读取各历史时刻与检测任务对应的解码结果，将各历史时刻与检测任务对应的解码结果对齐至当前时刻，得到当前时刻下各历史时刻检测任务对应的更新解码结果；

在一些实施例中，将各历史时刻检测任务对应的更新解码结果进行重构，生成各历史时刻对应的空间上对齐当前时刻的高维重构嵌入（Embedding）向量。

步骤S10212：基于所述检测任务对应的任务类型和第一历史时刻所述检测任务对应的重构特征，确定当前时刻所述检测任务对应的初始重构特征；所述第一历史时刻为当前时刻的上一时刻；

在一些实施例中，将第一历史时刻检测任务对应的重构特征进行传播（Propagation）用于初始化当前时刻初始重构特征，得到第一初始重构特征。

在一些实施例中，根据检测任务对应的任务类型，随机初始化生成多个第二初始重构特征。

在一些实施例中，将第一初始重构特征和第二初始重构特征进行拼接，得到当前时刻检测任务对应的初始重构特征。

在一些实施例中，运动目标检测任务对应的初始重构特征包括运动目标在BEV空间的初始位置及上下文特征；道路分割检测任务对应的初始重构特征包括BEV地平面 $S_H \times S_W$ 个同等大小矩形块面片对应的上下文特征，其位置固定为各个面片中心点；车道线检测任务对应的初始重构特征包括各条车道线或路沿的初始几何中心位置及上下文特征。

步骤S10213：将所述当前时刻所述检测任务对应的初始重构特征确定为所述第二查询信息；

步骤S10214：基于各第二历史时刻所述检测任务对应的重构特征和所述当前时刻所述检测任务对应的初始

重构特征, 确定所述第二键信息和所述第二值信息; 所述第二历史时刻为除所述第一历史时刻之外的历史时刻。

在一些实施例中, 将各第二历史时刻检测任务对应的重构特征和当前时刻检测任务对应的初始重构特征进行拼接得到第二键信息和第二值信息; 第二键信息和第二值信息一致。

本公开实施例中, 基于各历史时刻与当前时刻车辆的位姿转换关系, 对各历史时刻检测任务对应的检测结果进行重构, 确定各历史时刻检测任务对应的重构特征, 以及根据检测任务对应的任务类型和第一历史时刻检测任务对应的重构特征, 确定当前时刻检测任务对应的初始重构特征; 第一历史时刻为当前时刻的上一时刻; 之后, 将当前时刻检测任务对应的初始重构特征确定为第二查询信息; 之后, 基于各第二历史时刻检测任务对应的重构特征和当前时刻检测任务对应的初始重构特征, 确定第二键信息和第二值信息; 第二历史时刻为除第一历史时刻之外的历史时刻。这样, 第二查询信息包括当前时刻检测任务对应的初始重构特征, 如此, 可以在第一查询信息中融合当前时刻的任务相关信息和各历史时刻的任务相关信息, 进而可以在目标融合特征中建立当前时刻的任务相关信息和各历史时刻的任务相关信息之间的关联, 以提高环境检测的准确性。

在一些实施例中, 步骤S10211中所述的基于各所述历史时刻与当前时刻所述车辆的位姿转换关系, 对各所述历史时刻所述检测任务对应的检测结果进行重构, 确定各所述历史时刻所述检测任务对应的重构特征, 可以包括如下步骤S102111和步骤S102112, 其中:

步骤S102111: 基于所述位姿转换关系, 将各所述历史时刻所述检测任务对应的检测结果对齐至所述当前时刻, 确定各所述历史时刻所述检测任务对应的更新检测结果;

在一些实施例中, 根据位姿转换关系, 将各历史时刻存储的运动目标检测任务、道路分割检测任务和车道线检测任务对应的检测结果转换对齐至当前时刻本车坐标系下。

在一些实施例中, 由于道路分割检测任务及车道线检测任务对应的中心位置的维度为2, 需在高度Z轴方向上补充地面在本车坐标系下的高度值, 将其扩充为3维, 再根据转换公式计算各历史时刻对齐至当前时刻道路分割检测任务及车道线检测任务对应的中心位置的坐标。

步骤S102112: 基于所述位姿转换关系、各所述历史时刻与所述当前时刻之间的时间差、所述检测任务对应的运动速度, 对各所述历史时刻所述检测任务对应的更新检测结果进行重构, 确定各所述历史时刻所述检测任务对应的重构特征。

在一些实施例中, 利用位姿转换关系、时间差及目标运动速度, 根据运动感知正则化 (MLN) 方法, 计算MLN对应系数, 并利用MLN对应系数对各历史时刻缓存的各检测结果对齐至当前时刻Embedding进行重构。

本公开实施例中, 通过根据位姿转换关系, 将各历史时刻所述检测任务对应的检测结果对齐至当前时刻, 确定各历史时刻检测任务对应的更新检测结果; 根据位姿转换关系、各历史时刻与当前时刻之间的时间差、检测任务对应的运动速度, 对各历史时刻检测任务对应的更新检测结果进行重构, 确定各历史时刻检测任务对应的重

构特征。这样，将历史时刻检测任务对应的检测结果对齐至当前时刻并进行重构，以使当前时刻任务相关信息和历史时刻任务相关信息在语义上保持连续性和一致性。

在一些实施例中，步骤S10212中所述的基于所述检测任务对应的任务类型和第一历史时刻所述检测任务对应的重构特征，确定当前时刻所述检测任务对应的初始重构特征，可以包括如下步骤S102121至步骤S102123，其中：

步骤S102121：基于所述第一历史时刻所述检测任务对应的重构特征，确定所述当前时刻所述检测任务对应的第一初始重构特征；

在一些实施例中，将第一历史时刻检测任务对应的重构特征进行Propagation传播，得到当前时刻检测任务对应的第一初始重构特征。

步骤S102122：基于所述检测任务对应的任务类型，随机生成所述当前时刻所述检测任务对应的第二初始重构特征；

步骤S102123：将所述第一初始重构特征和所述第二初始重构特征进行拼接，确定所述当前时刻对应的初始重构特征。

在一些实施例中，运动目标检测任务和车道线检测任务直接将第一初始重构特征和第二初始重构特征进行拼接，得到初始重构特征。

在一些实施例中，道路分割检测任务将第一初始重构特征和第二初始重构特征进行拼接后，还需确认第一历史时刻重构的Embedding所对应在当前时刻本车坐标系下BEV空间的2D地平面上 $S_H \times S_W$ 个矩形块面片中的所属位置。

本公开实施例中，根据第一历史时刻检测任务对应的重构特征，确定当前时刻检测任务对应的第一初始重构特征；之后，根据检测任务对应的任务类型，随机生成当前时刻检测任务对应的第二初始重构特征；之后，将第一初始重构特征和第二初始重构特征进行拼接，确定当前时刻对应的初始重构特征。这样，一方面，利用第一历史时刻检测任务对应的重构特征初始化可以使当前时刻任务相关信息和历史时刻任务相关信息保持一定的连续性，有助于减少因随机初始化而导致的不稳定性，另一方面，利用随机初始化学到更多的多样性信息。

在一些实施例中，步骤S102122中所述的基于所述检测任务对应的任务类型，随机生成所述当前时刻所述检测任务对应的第二初始重构特征，可以包括如下步骤S1021221和步骤S1021222，其中：

步骤S1021221：基于所述检测任务对应的任务类型，随机生成当前时刻所述检测任务对应的初始特征；

在一些实施例中，根据不同的任务类型，生成检测任务对应的初始特征。

在一些实施例中，各检测任务的上下文Embedding向量初始值均设为0向量。

步骤S1021222：对所述检测任务对应的初始特征进行重构，确定当前时刻所述检测任务对应的第二初始重

构特征。

本公开实施例中，根据检测任务对应的任务类型，随机生成当前时刻检测任务对应的初始特征；对检测任务对应的初始特征进行重构，确定当前时刻检测任务对应的第二初始重构特征。这样，对检测任务对应的初始特征进行重构保持与各历史时刻特征的一致性。

在一些实施例中，步骤S1021221中所述的基于所述检测任务对应的任务类型，随机生成当前时刻所述检测任务对应的初始特征，可以包括如下步骤S1021221至步骤S1021223中的至少之一：

步骤S1611：在所述检测任务中包括运动目标检测任务的情况下，对三维空间的坐标进行均匀采样，确定所述运动目标检测任务对应的初始特征；

在一些实施例中，运动目标检测任务对应的中心位置初始化采用本车坐标系下BEV空间下3D空间中的位置均匀分布坐标。

步骤S1612：在所述检测任务中包括道路分割检测任务的情况下，基于设定的车辆可通行区域中各子区域的中心位置的坐标，确定所述道路分割检测任务对应的初始特征；

在一些实施例中，道路分割检测任务对应的中心位置初始化采用 $S_H \times S_W$ 个矩形块面片（即子区域）中第一历史时刻Propagation传播后剩余的面片中心位置坐标。

步骤S1613：在所述检测任务中包括车道线检测任务的情况下，对二维空间的坐标进行均匀采样，确定所述车道线检测任务对应的初始特征。

在一些实施例中，车道线检测任务中心位置初始化采用本车坐标系下BEV空间下2D地平面上的位置均匀分布坐标。

本公开实施例中，在检测任务中包括运动目标检测任务的情况下，对三维空间的坐标进行均匀采样，确定运动目标检测任务对应的初始特征；在检测任务中包括道路分割检测任务的情况下，基于设定的车辆可通行区域中各子区域的中心位置的坐标，确定道路分割检测任务对应的初始特征；在检测任务中包括车道线检测任务的情况下，对二维空间的坐标进行均匀采样，确定车道线检测任务对应的初始特征。这样，针对不同的检测任务采用不同的初始化方式，提高了初始化的准确性。

在一些实施例中，步骤S103中所述的基于所述周视图像特征和所述环视图像特征，确定第一键信息和第一值信息，可以包括如下步骤S1031和步骤S1032，其中：

步骤S1031：将所述二维周视特征和所述二维环视特征进行拼接，确定所述检测任务对应的第一键信息；

步骤S1032：将所述二维周视特征、所述三维周视位置特征、所述二维环视特征和所述三维环视位置特征进行拼接，确定所述检测任务对应的第一值信息。

本公开实施例中，将二维周视特征和所述二维环视特征进行拼接，确定检测任务对应的第一键信息；将二

维周视特征、三维周视位置特征、二维环视特征和三维环视位置特征进行拼接,确定检测任务对应的第一值信息。这样,可以使得第一值信息和第一键信息中包括完整的周视图像特征和环视图像特征,提高对车辆的周围环境检测的准确性。

在一些实施例中,步骤S104所述的利用注意力机制,对所述第一查询信息、所述第一键信息和所述第一值信息进行处理,得到目标融合特征,可以包括如下步骤S1041和步骤S1042,其中:

步骤S1041:将所述第一查询信息作为所述注意力机制的查询矩阵、所述第一键信息作为所述注意力机制的键矩阵,生成注意力矩阵;

步骤S1042:基于所述注意力矩阵和所述第一值信息,确定所述目标融合特征。

本公开实施例中,将第一查询信息作为注意力机制的查询矩阵、第一键信息作为注意力机制的键矩阵,生成注意力矩阵;基于注意力矩阵和第一值信息,确定目标融合特征。这样,一方面,利用注意力机制对任务相关信息、周视图像特征、环视图像特征进行稀疏化表达,从而减少了计算量和内存占用;另一方面,利用注意力机制将周视图像特征和环视图像特征进行融合,以及建立各历史时刻与当前时刻之间的关联,以提高对车辆的周围环境检测的准确性。

在一些实施例中,所述检测任务的数量为多个;步骤S102中所述的基于检测任务对应的任务类型,确定第一查询信息,可以包括如下步骤S1023和步骤S1024,其中:

步骤S1023:针对每一所述检测任务,基于所述检测任务对应的任务类型,确定所述检测任务对应的查询信息;

在一些实施例中,检测任务可以包括运动目标检测任务、道路分割检测任务、车道线检测任务等中的至少之一。

在一些实施例中,每一检测任务对应不同的查询信息。

步骤S1024:将各所述检测任务对应的查询信息进行拼接,确定所述第一查询信息。

本公开实施例中,针对每一检测任务,基于检测任务对应的任务类型,确定检测任务对应的查询信息;将各检测任务对应的查询信息进行拼接,确定第一查询信息。这样,多任务融合实现对道路环境的全面感知,提高对车辆周围环境检测的准确性和鲁棒性。

在一些实施例中,上述方法还可以包括如下步骤S106,其中:

步骤S106:对所述车辆的雷达点云数据进行特征提取,得到点云特征;

在一些实施例中,通过激光雷达扫描车辆周围环境并收集反射信号,得到雷达点云数据,雷达点云数据提供了车辆周围环境的高精度三维感知信息,能够精确反映物体与环境的3D空间深度位置,对空间物体测距与精确定位拥有视觉无可比拟的优势,例如对于突发闯入或非白名单上的未知障碍物避障更是具有优于视觉的探测

能力。

上述步骤S103中所述的基于所述周视图像特征和所述环视图像特征，确定第一键信息和第一值信息，可以包括如下步骤S1033，其中：

步骤S1033：基于所述周视图像特征、所述环视图像特征和所述点云特征，确定第一键信息和第一值信息。

在一些实施例中，将纯视觉的周视图像特征和环视图像特征与点云特征进行融合，以提高对车辆作为环境检测的准确度，即可以将周视图像特征、环视图像特征和点云特征进行拼接，得到第一键信息和第一值信息。

在一些实施例中，将二维周视特征、二维环视特征和点云特征进行拼接，确定检测任务对应的第一键信息；将二维周视特征、三维周视位置特征、二维环视特征、三维环视位置特征和点云特征进行拼接，确定检测任务对应的第一值信息。

本公开实施例中，通过对车辆的雷达点云数据进行特征提取，得到点云特征，并根据周视图像特征、环视图像特征和点云特征，确定第一键信息和第一值信息，以使第一键信息和第一值信息中包括纯视觉的周视图像特征和环视图像特征、以及点云特征，提高对车辆的周围环境检测的准确性。

下面说明本公开实施例在实际场景中的应用。

随着自动驾驶技术的快速发展，当前基于多源多模态BEV融合的感知方法已经成为行、泊车感知主流技术方式，受到行业普遍青睐。BEV感知利用车辆周身相机及雷达等多源多模态信息，通过将各传感器特征信息统一向BEV空间映射，实现车辆前后左右远中近距全方位特征信息融合与覆盖，从而大幅提升对外部重要目标和整体道路拓扑的感知能力。这种不同于单视角的感知方式是在各传感器数据的特征层直接进行不同源、不同模态的空间覆盖式信息融合，不仅可以避免单传感器截断目标干扰以及后融合带来的有用信息丢失，而且可以大幅增强空间特征表达能力，进而有效提升全局环境感知探测的准确性。

在多模态BEV感知中，激光雷达能够精确反映物体与环境的3D空间深度位置，对空间物体测距与精确定位拥有视觉无可比拟的优势，例如对于突发闯入或非白名单上的未知障碍物避障更是具有优于视觉的探测能力。然而，激光雷达居高不下的高昂成本较大限制了以雷摄融合为主的BEV感知方式的走量推广，使其当前局限于适配较高端车型，难以大规模下探打入到中低端车型市场。鉴于此，以纯视觉为代表的低成本传感融合解决方案对于行业多数企业来说更具性价比，为大规模市场的规模走量提供了可能性。视觉图像信息具有细节丰富、特征饱满的特点，足以充分反映车辆所处周边环境场景及针对性目标的色彩、形状、纹理、位姿等多元信息，始终被视为分类识别与场景语义表征的较优数源。然而，仅依靠视觉图像内容的缺点同样突出，即其所反映的只有2D平面信息，不似激光点云自然拥有良好的景深属性，注定了它难以实现高精度的测量与定位。对此，尽管当前学术界利用深度神经网络已经提出了多种方法用以提升视觉景深的预测精度，但现阶段针对图像的景深预测仍显粗粒度，如何实现对2D图像中目标在3D物理世界的精准定位仍构成该研究领域的重大挑战。

针对上述问题,在不考虑激光雷达传感器的情况下,一种直观的缓解方式是尽可能充分利用车载各视角图像信息来提升对目标和环境的综合感知定位能力。在行、泊车功能需求的前提支持下,自动驾驶车辆均会配备足以覆盖车周360°视角的周视与环视相机,实现对车周例如中近距离的探测覆盖。然而,在进行量产应用方面,由于周视图像与环视图像在视野范围(Field of View, FOV)、畸变模式等方面存在较大差异,周视与环视针对行、泊任务一般是各自独立处理,在特征层面真正实现融合存在一定的技术难度,通常在后处理阶段对二者的输出信息进行融合利用。这种方法带来了两方面问题,一方面是两套独立模型会增加车端有限的资源消耗且因抢占造成的高延时,另一方面则是对周视固有的视野盲区特征缺陷无法有效弥补。因此,可以构建一种有效的一体化网络结构,充分地将周视与环视图像信息高效融合形成真正的多视多源感知输入输出,不仅可以显著解决前述问题,而且还能带来感知定位准确率的提升。

自然语言处理和计算机视觉领域的大量研究实践表明,相比视觉感受野受限的CNN结构,基于注意力(Attention)机制的神经网络架构(Transformer网络架构)在构建全局与局部特征关联、同源与非同源任务关联等方面更具优势,能够大幅提升对复杂场景的预测判断能力。基于这一特点,充分利用Transformer架构优势,通过Attention机制和高效率低开销的稀疏化Query表达,将BEV感知网络中预测对象与周视、环视图像信息、动态目标与静态道路结构信息、当前时刻与历史前后信息的抽象关联关系建立起来,实现跨全视角相机和跨全任务内容的充分关联计算,有助于在确保运行效率的同时显著提升模型对整体复杂环境的感知预测准确性,具有较好的实施可行性。

针孔相机模型下的多路周视图像BEV感知任务,只对应输出3D目标信息,无法实现对环视鱼眼相机图像的融合利用,且不支持道路信息等多任务输出,更没有考虑动静目标之间的关联关系,使用场景比较局限。

另外,针对多路鱼眼相机获取若干图像,对当前帧鱼眼图像以及历史帧鱼眼图像进行特征提取,得到当前帧鱼眼图像对应的第一特征图以及历史帧鱼眼图像对应的第二特征图,再根据第一特征图以及第二特征图获取BEV合并特征,之后根据BEV合并特征进行目标跟踪检测。该方法仅针对鱼眼图像进行特征提取与映射处理,鱼眼畸变对识别和精确定位不可避免地会造成较大负向影响,且鱼眼环视探测距离受限,同时,模型无法复用常规针孔相机图像,应用场景也十分有限。

基于Transformer结构的稀疏Query设计,以动态目标对象为中心的时间机制,模型以在线方式执行,长期历史缓存信息通过逐帧的对象查询传播,还引入了一种运动感知层归一化来对物体的运动进行建模。尽管该方法也是基于Transformer构建了一套稀疏化BEV感知架构,但只考虑了传统的周视图像输入,没有对环视鱼眼图像信息的融合机制,输出任务也只有特定3D动态目标,不具备对道路静态结构的感知能力,自然也无从构建周视、环视、动态、静态等重要环境要素之间的关联关系,难以应对复杂多任务场景。

基于以上描述,本公开实施例提供了一种稀疏化HAS-Propagation Transformer(Hierarchical Attention Structure

Propagation Transformer, 分层注意力结构传播转换器) 架构的BEV周环视动静态多任务感知方法, 通过构建一套基于时序稀疏化特征表达的Transformer模型架构, 实现对覆盖车周 360° 范围的周视、环视图像一体化感知多任务处理输出。这种Transformer模型架构依靠车辆周身视觉传感器输入图像信息来解决行泊车过程中遇到的动静态目标与环境感知问题, 降低探测传感器成本。周视与环视的一体化输入输出结构, 一方面, 避免了传统周视、环视各自感知任务独立运行带来的资源抢占与延时等待, 有效提升信息处理效率; 另一方面, 通过将周视、环视信息融合利用以大幅提升感知准确率。该Transformer架构采用的动静态感知多任务稀疏化特征表达, 不仅可以有效改善现今广泛应用的密集型BEV特征映射生成所造成的大量内存消耗与延时降效, 而且HAS-Propagation Transformer结构中的Hybrid Attention与AS Attention机制实现了动态目标与静态道路结构之间、道路可通行区域与路沿车道线之间、当前状态与历史状态之间的抽象特征关联, 从而使输出动态目标与静态道路环境之间的位置拓扑关系更加精准。总言之, 本公开实施例全面利用了自车低成本视觉传感器信息, 基于Transformer构建了时序动静态目标有效关联关系, 以实现输入输出一体化高效感知输出。

图2示出了本公开实施例提出的一种利用稀疏化HAS-Propagation Transformer架构进行车辆环境检测的结构示意图, 如图2所示的结构包括推理部200和训练部240, 其中, 推理部200中各部如下所示: 在周视针孔摄像机成像部201获取周视图像, 在周视编码网络2D图像特征编码部202进行周视图像2D编码以及在周视成像3D伪点云部203和周视3D位置编码部204完成对周视图像的3D编码; 在环视鱼眼摄像机成像部205获取环视图像, 在环视编码网络2D图像编码部206进行环视图像2D编码以及在环视成像3D伪点云207和环视3D位置编码部208完成对环视图像的3D编码; 根据车身定位姿态及时间戳信息214在历史缓存信息部212缓存历史时刻的信息, 以及根据历史缓存信息在历史Embedding重构生成部211生成历史Embedding重构向量; 利用历史Embedding重构向量在当前Query初始化部213初始化当前Query; 将上述各部获取的周视图像2D编码特征、3D编码特征、环视图像2D编码特征、3D编码特征、历史Embedding重构向量、当前Query初始化输入到HAS-Propagation Transformer网络部220中, 分别在3D目标检测任务解码部231、道路分割检测任务解码部232和车道线检测任务解码部233对HAS-Propagation Transformer网络部220输出的特征进行解码。训练部240中各部如下所示: 利用初始Query优化迭代241对当前Query初始化部213进行训练; 利用3D目标标注真值信息242对3D目标检测任务解码部231进行训练; 利用道路结构化标注真值信息243对道路分割检测任务解码部232和车道线检测任务解码部233进行训练。

图3示出了本公开实施例提出的一种利用稀疏化HAS-Propagation Transformer架构进行车辆环境检测的网络结构示意图, 如图3所示, 利用周视编码网络301对当前帧周视图像 $1 \sim N_a$ 进行2D编码, 得到周视特征 F_{2d}^a , 利用周视3D位置编码302对周视图像进行3D编码, 得到周视3D编码 F_{3d}^a ; 利用环视编码网络303对当前帧环视图像 $1 \sim N_s$ 进行2D编码, 得到环视特征 F_{2d}^s , 利用环视3D位置编码304对环视图像进行3D编码, 得到环视3D编码 F_{3d}^s ; 根据历史缓存队列305中的上一帧时刻Embedding重构向量307初始化, 得到当前帧初始化Query信息308;

将历史缓存队列305中的信息与当前帧时空对齐, 并进行重构, 得到历史帧Embedding重构向量306; 将周视特征 F_{2d}^a 、周视3D编码 F_{3d}^a 、环视特征 F_{2d}^s 、环视3D编码 F_{3d}^s 、当前帧初始化Query信息308和历史帧Embedding重构向量306输入到HAS-Propagation Transformer网络部220中, 输出3D目标检测任务Query309、道路分割检测任务Query310、车道线检测任务Query311, 并利用3D目标检测解码任务头312、道路分割检测解码任务头313、车道线检测解码任务头314分别对3D目标检测任务Query309、道路分割检测任务Query310、车道线检测任务Query311进行解码, 得到当前帧 $Top-K$ 个缓存315; 将当前帧 $Top-K$ 个缓存315存入历史缓存队列305中。

本公开实施例中的稀疏化HAS-Propagation Transformer架构的BEV周环视动静态多任务感知方法的实现过程如图4所示, 可以包括步骤S401至步骤S406:

步骤S401: 周视、环视图像: 2D特征编码和3D位置编码;

本公开实施例提供的坐标系如图5所示, 包括相机坐标系501和本车坐标系502。

首先对周视针孔透视相机图像和环视鱼眼相机图像, 分别利用两个独立的2D卷积编码主干网络进行特征编码, 获取周视图像2D编码特征 F_{2d}^a 和环视图像2D编码特征 F_{2d}^s ; 然后, 分别利用周视相机内参及其畸变参数模型, 环视相机内参及其畸变参数模型, 构建具有一定深度间隔的各相机图像去畸变后的视锥伪点云, 并结合对应相机外参将其坐标从相机坐标系501转换至本车坐标系502下, 完成周视图像中各像素点对应的3D位置编码 F_{3d}^a 和环视图像中各像素点对应的3D位置编码 F_{3d}^s 。

在一些实施例中, 上述步骤S401可以包括以下步骤S4011和S4012:

步骤S4011: 对周视相机图像和环视相机图像进行2D特征编码 F_{2d}^a 、 F_{2d}^s , 主干网络通常可采用但不限于具有一定深度层级的如ResNet、EfficientNet、SwinTransformer、VoVNetV2等系列常用2D卷积网络结构, 且所有周视图像和所有环视图像分别各自共享一套网络权重。

步骤S4012: 对周视相机图像和环视相机图像进行3D位置编码 F_{3d}^a 、 F_{3d}^s , 其操作方式为, 针对某视角相机图像:

首先, 根据该周视或环视相机的畸变模型及畸变参数, 结合对应相机内参 $I_c \in \mathbb{R}^{3 \times 3}$, 计算还原图像中各像素点在相机坐标系下对应的去畸变真实矢量方向。

接着, 根据所需生成的BEV探测物理空间大小 $H_B \times W_B$, 设置合理的等间隔深度区段 $\{z_k | k \in n\}$, 并结合各像素点的真实矢量方向, 构建图像中各像素点 (i, j) 在相机坐标系下不同深度区段的坐标值 $(x_{i,j,k}^c, y_{i,j,k}^c, z_{i,j,k}^c)$ 。

然后, 利用对应相机相对本车坐标系下的外参旋转 $R_c \in \mathbb{R}^{3 \times 3}$ 和位移 $T_c \in \mathbb{R}^{3 \times 1}$ 信息, 将各像素点在不同深度区段下的坐标转换至本车坐标系下 $(x_{i,j,k}^e, y_{i,j,k}^e, z_{i,j,k}^e)$ 。

之后, 采用RefPE编码对本车坐标系下各像素点对应的不同深度区段坐标进行编码, 对于坐标点

$(x_{i,j,k}^e, y_{i,j,k}^e, z_{i,j,k}^e)$, 其编码方式参见公式 (1) 和公式 (2):

$$\psi_{i,j,k} = \text{Concat} \left\{ \xi(d_{i,j,k}), \xi(\sin \theta_{i,j,k}), \xi(\cos \theta_{i,j,k}), \xi(z_{i,j,k}^e) \right\} \quad (1);$$

$$\psi_{i,j} = \text{MLP} \left(\sum_{k=1}^n s_{i,j,k} \times \psi_{i,j,k} \right) \quad (2);$$

$$\text{其中, } d_{i,j,k} = \sqrt{x_{i,j,k}^e{}^2 + y_{i,j,k}^e{}^2}, \quad \sin \theta_{i,j,k} = \frac{y_{i,j,k}^e}{d_{i,j,k}}, \quad \cos \theta_{i,j,k} = \frac{x_{i,j,k}^e}{d_{i,j,k}}, \quad \xi \text{ 表示傅里叶位置}$$

编码, $s_{i,j,k}$ 表示像素点 (i, j) 对应的深度区段 k 的加权重, 可以人工设定或训练获得, Concat 和 MLP 则分别表示拼接操作和多层感知机网络。

步骤S402: 历史缓存信息Embedding重构生成;

在一些实施例中, 从历史缓存Memory中读取 $t-1$ 、 $t-2$ 至 $t-n$ 帧时刻存储的关于3D目标检测任务 (对应前述实施例中的运动目标检测任务)、道路可通行区域分割检测任务 (对应前述实施例中的道路分割检测任务)、车道线路沿检测任务 (对应前述实施例中的车道线检测任务) 等动静态各任务表达缓存信息, 并结合各历史帧时刻 ($t-1$ 、 $t-2$ $t-n$) 与当前帧 t 时刻之间的位姿转换关系, 对各检测结果表达进行重构, 生成对应的空间上对齐当前帧 t 时刻的高维重构Embedding向量。

在一些实施例中, 历史缓存Memory中的 $t-1$ 、 $t-2$ 至 $t-n$ 帧时刻如图6所示, FIFO缓存600中缓存 $t-1$ 、 $t-2$ 至 $t-n$ 帧时刻的信息, 其中, $t-1$ 帧缓存信息601包括 $t-1$ 帧全局信息604、 $TopK_d$ 个 $t-1$ 帧3D目标检测任务信息605、 $TopK_s$ 个 $t-1$ 帧道路分割检测任务信息606、 $TopK_l$ 个 $t-1$ 帧车道线路沿检测任务信息607, $t-1$ 帧全局信息604包括 $t-1$ 帧位姿信息602和 $t-1$ 帧时间戳信息603; $t-k$ 帧缓存信息611包括 $t-k$ 帧全局信息614、 $TopK_d$ 个 $t-k$ 帧3D目标检测任务信息615、 $TopK_s$ 个 $t-k$ 帧道路分割检测任务信息616、 $TopK_l$ 个 $t-k$ 帧车道线路沿检测任务信息617, $t-k$ 帧全局信息614包括 $t-k$ 帧位姿信息612和 $t-k$ 帧时间戳信息613。可选地, 以 $t-k$ 时刻为例, 针对该时刻存储的信息包括: $t-k$ 时刻对应的本车相对于全局坐标系位姿 E_{t-k} 及时间戳 T_{t-k} ; $TopK_d$ 个3D目标检测任务对应的上下文Embedding向量 $Q_{c,(t-k)}^{det}$ 、中心位置 $P_{(t-k)}^{det}$ 、目标运动速度 $V_{(t-k)}^{det}$; $TopK_s$ 个道路分割检测任务对应的上下文Embedding向量 $Q_{c,(t-k)}^{seg}$ 、中心位置 $P_{(t-k)}^{seg}$; $TopK_l$ 个车道线路沿检测任务对应的上下文Embedding向量 $Q_{c,(t-k)}^{lane}$ 、中心位置 $P_{(t-k)}^{lane}$ 。通常情况下, 考虑各类任务上下文Embedding向量的维度均为 C , 3D目标中心位置和速度的维度分别为3和2, 道路及车道线中心位置的维度为2, 以及该时刻时间戳及本车位姿信息的维度为 $3 \times 3 + 3 \times 1 + 1$, 缓存Memory的总占用为 $N \times (13 + K_d \times (C + 5) + K_s \times (C + 2) + K_l \times (C + 2))$ 。

在一些实施例中, 结合各历史帧时刻 ($t-1$ 、 $t-2$ $t-n$) 与当前帧 t 时刻之间的位姿转换关系, 对各任务信息表达进行重构, 以历史 $t-k$ 时刻为例, 生成 $t-k$ 时刻各检测任务对应的空间上对齐当前帧 t 时刻的高维重构Embedding向量, 如图7、图8、图9所示, 操作方式如下:

首先, 计算 $t-k$ 时刻与 t 时刻的时间差 $\Delta T = T_t - T_{t-k}$, 及 $t-k$ 时刻与 t 时刻的本车位姿转换关系 $E_{t-k}^t = E_t^{-1} \cdot E_{t-k}$ 并根据 E_{t-k}^t 对 $t-k$ 时刻存储的3D目标检测任务、道路分割检测任务、车道线路沿检测任务中心位置转换对齐至当前时刻本车坐标系下, 由于道路及车道线中心位置的维度为2, 需在高度Z轴方向上补充地面在本车坐标系下的高度值, 将其扩充为3维, 再根据转换公式计算 $P_{(t-k) \rightarrow t} = E_{t-k}^t \cdot P_{(t-k)}$ 。

然后, 利用本车位姿转换关系 E_{t-k}^t 时间差 ΔT 及目标运动速度 $V_{(t-k)}$, 根据运动感知正则化 (MLN) 方法, 计算MLN对应系数 (γ, β) 如下:

$$\begin{cases} \gamma = \xi_{\gamma}(E_{t-k}^t, \Delta T, V_{(t-k)}) \\ \beta = \xi_{\beta}(E_{t-k}^t, \Delta T, V_{(t-k)}) \end{cases} \quad (3);$$

其中, ξ_{γ} 和 ξ_{β} 分别表示两个独立的线性变换层, 网络权重由学习获得, 目标运动速度对于3D目标 $V_{(t-k)} = V_{(t-k)}^{det}$, 对于道路和车道线目标运动速度 $V_{(t-k)} = \mathbf{0}$, 也就是说, 道路可通行区域目标和车道线目标具有相同的系数 (γ, β) , 而对于3D运动目标, 由于它具有不同的速度, 获得的系数 (γ, β) 也有所差异。

之后, 结合前述获得的MLN系数 (γ, β) , $t-k$ 时刻缓存的各任务信息对齐 t 时刻Embedding进行重构, 计算方式如下:

$$\widetilde{Q_{c,(t-k)}} = \gamma \cdot \text{LN}(Q_{c,(t-k)}) + \beta \quad (4);$$

$$\widetilde{Q_{p,(t-k)}} = \gamma \cdot \text{LN}(\text{MLP}(P_{(t-k) \rightarrow t})) + \beta \quad (5);$$

$$\widetilde{Q_{(t-k)}} = \widetilde{Q_{c,(t-k)}} + \widetilde{Q_{p,(t-k)}} \quad (6);$$

其中, MLP表示多层感知机网络, LN表示层正则化。由于道路分割检测所对应的物理对象为BEV空间地平面上的各个方块面片中心位置, 直接利用 E_{t-k}^t 对 $P_{(t-k)}^{seg}$ 进行转换后得到的 $P_{(t-k) \rightarrow t}^{seg}$ 一般与当前 t 时刻本车坐标系下的BEV空间各个方块面片难以重合对应, 因此可以采用最近面片原则, 即将与 $P_{(t-k) \rightarrow t}^{seg}$ 位置最近的面片中心位置作为 $P_{(t-k) \rightarrow t}^{seg}$ 的值, 再带入到Embedding重构计算。

步骤S403: 当前帧Query构建与初始化;

在一些实施例中, 根据距离当前帧 t 时刻最近的上一帧 $t-1$ 时刻重构Embedding向量, 将其进行Propagation传播用于初始化当前帧Query信息, 同时结合部分当前帧空间位置随机初始化, 共同完成对包括3D目标、道路可通行区域、车道线路沿等动静各任务在内的初始Query构建, 特征向量内容包括所有Query对应的上下文Embedding和位置Embedding。

在一些实施例中, 初始化当前帧Query可以包括3D目标检测、道路可通行区域分割、车道线路沿检测等动静

态各任务对应的Query构建，其物理对象含义如图10所示，其中：BEV空间1001中，包括本车中心1002、3D目标检测任务Query309、道路分割检测任务Query310和车道线检测任务Query311，3D目标检测任务Query309对应为各个候选3D目标在BEV空间的初始位置及上下文特征；道路分割检测任务Query310对应为BEV地平面 $S_H \times S_W$ 个同等大小矩形块面片对应的上下文特征，其位置固定为各个面片中心点；车道线检测任务Query311对应为各条车道线或路沿的初始几何中心位置及上下文特征。

在一些实施例中，当前帧动静态各任务的Query初始化信息包括两个部分，第一部分是将上一帧 $t-1$ 时刻重构Embedding向量1001Propagation传播用于初始化当前帧Query信息1002，即： $\{\widetilde{Q}_{(t-1)} | \widetilde{Q}_{(t-1)} \in \widetilde{Q}_{(t-1)}^{\text{det}}, \widetilde{Q}_{(t-1)}^{\text{seg}}, \widetilde{Q}_{(t-1)}^{\text{lane}}\}$ ；第二部分是随机初始化一批当前帧 t 时刻Query，即： $\{\widetilde{Q}_{(t)} | \widetilde{Q}_{(t)} \in \widetilde{Q}_{(t)}^{\text{det}}, \widetilde{Q}_{(t)}^{\text{seg}}, \widetilde{Q}_{(t)}^{\text{lane}}\}$ 。

在一些实施例中，第一部分由上一帧 $t-1$ 时刻重构Embedding向量Propagation传播初始化当前帧Query中，3D目标任务的 $TopK_d$ 个 $\widetilde{Q}_{(t-1)}^{\text{det}}$ 和车道线路沿任务的 $TopK_l$ 个 $\widetilde{Q}_{(t-1)}^{\text{lane}}$ 均直接与第二部分对应任务Query拼接使用，而道路任务的 $TopK_s$ 个 $\widetilde{Q}_{(t-1)}^{\text{seg}}$ 除与第二部分对应任务Query拼接使用外，还可以确认这些 $t-1$ 时刻重构的Embedding所对应在当前帧本车坐标系下BEV空间2D地平面 $S_H \times S_W$ 个矩形块面片中的所属位置，即 $S_H \times S_W$ 个道路任务Query中的 $TopK_s$ 个由Propagation传播来初始化。

在一些实施例中，第二部分随机初始化的操作方式为：各3D目标任务Query的中心位置初始化采用本车坐标系下BEV空间3D空间中的位置均匀分布坐标；各道路任务Query的中心位置初始化采用步骤S433所述 $S_H \times S_W$ 个矩形块面片中上一帧Propagation传播后剩余的 $S_H \times S_W$ 中 $TopK_s$ 个面片中心位置坐标；各车道线路沿Query的中心位置为本车坐标系下BEV空间2D地平面上的位置均匀分布坐标；各任务的上下文Embedding向量初始值均设为0向量；为保持特征一致性，将前述获得的各任务Query位置坐标向量及上下文初始化Embedding向量代入公式(3)至公式(6)，得到第二部分初始化Embedding，即： $\{\widetilde{Q}_{(t)} | \widetilde{Q}_{(t)} \in \widetilde{Q}_{(t)}^{\text{det}}, \widetilde{Q}_{(t)}^{\text{seg}}, \widetilde{Q}_{(t)}^{\text{lane}}\}$ ，其中公式(3)中可设置 $(E_{t-t}^t = I, V_{(t-t)} = 0, \Delta T = 0)$ 代入计算MLN系数 (γ, β) 。

步骤S404：HAS-Propagation Transformer计算；

如图11所示，根据步骤S401、S402、S403所获得的信息均作为HAS-Propagation Transformer网络部的输入，该部包含若干层混合注意力机制1102 (Hybrid Attention) 与周视环视注意力机制1101 (AS Attention) 串联组合。其各对应输入依次为：根据步骤S402所获得的历史帧 $(t-1, t-2, \dots, t-n)$ 时刻动静态各任务Embedding和步骤S403所获得的当前帧 t 时刻初始化Query信息，将二者信息进行拼接，形成Hybrid Attention的输入Key向量和Value向量，并通过Attention机制建立动态目标与静态道路结构之间、道路可通行区域与路沿车道线之间、当前

状态与历史状态之间隐含的特征关联关系；然后，Hybrid Attention的输出作为后续AS Attention的Query输入，并根据步骤S401所获得的周视、环视图像2D编码特征 F_{2d}^a 、 F_{2d}^s 和3D位置编码特征 F_{3d}^a 、 F_{3d}^s ，形成AS Attention的输入Key向量和Value向量，再通过Attention机制建立动静态各任务与当前帧t时刻周视、环视图像观测内容之间的特征关联关系。这样，经过数层Hybrid Attention部与AS Attention部计算，当前帧Query完成了与各动静态任务之间和与各历史时刻之间的信息关联与更新表达。

在一些实施例中，Hybrid Attention部主要针对当前时刻各任务与对应历史任务之间、以及动静态各任务之间的关联关系构建，而AS Attention部主要针对动静态各任务与当前周视、环视观测之间的关联关系构建，Hybrid Attention和AS Attention均采用标准Transformer中的多头Attention计算方式，公式如下：

$$\left\{ \begin{array}{l} MultiHead(Q, K, V) = Concat(H_1, H_2, \dots, H_m) \times W^O \\ H_i = Attention(Q \times W_i^Q, K \times W_i^K, V \times W_i^V) \\ Attention(Q, K, V) = softmax\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V \end{array} \right. \quad (7);$$

其中，Q、K、V分别表示Attention部输入Query（对应前述实施例中的查询矩阵）、Key（对应前述实施例中的键矩阵）、Value（对应前述实施例中的值矩阵）， d_k 表示K向量的维度， W_i^Q 、 W_i^K 、 W_i^V 、 W^O 分别为多头Attention的权系数。

在一些实施例中，Hybrid Attention部输入输出如图11所示，可以包括如下操作：

设置Query由上一帧 $t-1$ 时刻重构Embedding向量Propagation传播至当前帧t时刻 $\{\widetilde{Q}_{(t-1)} | \widetilde{Q}_{(t-1)} \in \widetilde{Q}_{(t-1)}^{det}, \widetilde{Q}_{(t-1)}^{seg}, \widetilde{Q}_{(t-1)}^{lane}\}$ 和一批当前帧t时刻的随机初始化 $\{\widetilde{Q}_{(t)} | \widetilde{Q}_{(t)} \in \widetilde{Q}_{(t)}^{det}, \widetilde{Q}_{(t)}^{seg}, \widetilde{Q}_{(t)}^{lane}\}$ 共同拼接组成，记为 $\{Q_{hyb}^{(in)} | Q_{hyb}^{(in)} \in Concat(\widetilde{Q}_{(t)}, \widetilde{Q}_{(t-1)})\}$ 。

然后，设置Key和Value完全相同，均为历史缓存除去 $t-1$ 时刻后剩余其他时刻 ($t-2, \dots, t-n$) 对齐当前帧t时刻的高维重构Embedding向量 $\{Q_{mem} | Q_{mem} \in \widetilde{Q}_{(t-2)}, \dots, \widetilde{Q}_{(t-n)}\}$ 与前述构建的当前初始化 $\{Q_{hyb}^{(in)} | Q_{hyb}^{(in)} \in Concat(\widetilde{Q}_{(t)}, \widetilde{Q}_{(t-1)})\}$ 共同拼接组成，即为 $\{K_{hyb} | K_{hyb} \in Concat(Q_{hyb}^{(in)}, Q_{mem})\}$ 和 $\{V_{hyb} | V_{hyb} \in Concat(Q_{hyb}^{(in)}, Q_{mem})\}$ 。

之后，将前述构建的Query、Key、Value代入多头Attention计算公式 (7)，获得Hybrid Attention部输出 $\{Q_{hyb}^{(out)}\}$ ，作为后续AS Attention部的输入使用。

在一些实施例中，AS Attention部输入输出如图11所示，可以包括如下操作：

Query设置为Hybrid Attention部的输出, 即 $\{Q_{AS}^{(in)} | Q_{AS}^{(in)} \in Q_{hyb}^{(out)}\}$;

然后, Value设置为所述步骤S401中的所有周视图像2D编码特征 F_{2d}^a 和环视图像2D编码特征 F_{2d}^s 拼接组合, 即为 $\{V_{AS} | V_{AS} \in \text{Concat}(F_{2d}^a, F_{2d}^s)\}$;

之后, Key设置为步骤S301中的所有周视图像2D编码特征 F_{2d}^a 和3D位置特征 F_{3d}^a 加和、与环视图像2D编码特征 F_{2d}^s 和3D位置特征 F_{3d}^s 加和, 共同拼接组合, 即为 $\{K_{AS} | K_{AS} \in \text{Concat}(F_{2d}^a + F_{3d}^a, F_{2d}^s + F_{3d}^s)\}$;

之后, 将前述构建的Query、Key、Value代入多头Attention计算公式(7), 获得AS Attention部输出 $\{Q_{AS}^{(out)}\}$, 如该层不是最后层, $\{Q_{AS}^{(out)}\}$ 作为下一层HAS-Propagation Transformer网络部的输入使用, 如该层是最后层, $\{Q_{AS}^{(out)}\}$ 则作为3D目标检测、道路可通行区域分割及车道线路沿检测解码使用。

步骤S405: 动静态多任务解码;

在一些实施例中, 根据步骤S404所获得的HAS-Propagation Transformer网络部输出的3D目标Query、道路可行驶区域Query和车道线路沿Query, 分别接入各任务对应解码头网络, 完成动静态目标与环境感知多任务输出; 然后, 对各任务Query的解码置信度由高到低进行降序排序, 分别选择排名靠前的 $TopK_d$ 个3D目标、 $TopK_s$ 个道路分割和 $TopK_l$ 个车道线路沿Query对应相关解码信息, 采用先进先出(FIFO)方式对历史缓存Memory进行更新, 供后一帧处理使用。

在一些实施例中, 分别对HAS-Propagation Transformer输出的3D目标Query、道路可行驶区域Query和车道线路沿Query进行解码, 其中: 3D目标Query的解码可采用简单的多层感知机网络, 输入维度为C, 输出维度则对齐目标置信度、目标类别、中心坐标、矩形框长宽高、目标朝向及目标速度等信息量, 可以根据实际需要设定; 道路可行驶区域Query的解码, 可以分割出BEV平面上的像素级位置区域, 由于每个道路Query对应BEV平面 $S_H \times S_w$ 个矩形块面片中的一个, 每个面片对应 $r_h \times r_w$ 像素大小, 因此可采用输入维度是C、输出维度是 $r_h \times r_w$ 加1维置信度的简单多层感知机网络, 当然, 此处也可采用简单的CNN网络对 $r_h \times r_w$ 进行像素分割; 车道线路沿Query也与前述方法一样是采用多层感知机网络, 输入维度是C, 但其输出维度取决于对车道线路沿的矢量化表达形式, 通常较为典型的有MapTR、PETRv2等表达方式, 甚至可采用3D车道线PersFormer方式(如采用3D车道线, 前述涉及的Query初始化中心点高度坐标可以做适配调整), 可结合实际需要及复杂度限制进行衡量, 此处不做约束。

在一些实施例中, 对3D目标检测、道路可通行区域分割及车道线路沿检测各任务Query的解码置信度由高到低进行降序排序, 分别选择排名靠前的 $TopK_d$ 个3D目标、 $TopK_s$ 个道路分割和 $TopK_l$ 个车道线路沿Query

对应的解码信息,然后根据各任务存入历史缓存Memory数据信息,采用先进先出(FIFO)方式对历史缓存Memory进行更新,供后一帧处理使用。

在一些实施例中,存入历史缓存Memory中的当前帧车道线路沿中心位置 $P_{(t)}^{lane}$, 其计算方式为:首先提取该车道线或路沿解码后的所有 k 个关键点坐标 $\{p_1, p_2, \dots, p_k\}$, 将 C 维车道线路沿Query作为输入, 利用多层感知机网络得到对应于各关键点的 k 维权重 $[w_1, w_2, \dots, w_k]$, 然后根据下述公式计算获得车道线路沿加权几何中心:

$$P_{(t)}^{lane} = \frac{\sum_{i=1}^k w_i \times p_i}{\sum_{i=1}^k w_i} \quad (8)$$

其中, 为保持多层感知机网络权重分布的有效性, 代入公式(8)的 k 个关键点坐标 $\{p_1, p_2, \dots, p_k\}$ 会事先根据解码信息进行一定规则逻辑的排序。

步骤S406: 样本库制备、损失函数设计、模型训练及推理。

在一些实施例中, 对步骤S401至步骤S405所述的BEV感知网络模型进行监督训练, 采集获取各时刻对应的车载各视角周视针孔摄像机和环视鱼眼摄像机2D视觉图像, 实现传感器时间戳数据对齐配对; 然后, 针对不同的特定感知任务, 完成相应的任务数据样本标注, 生成网络模型监督训练所需真值; 之后, 通过构造相应多任务损失函数, 结合有效的加快训练收敛策略, 完成BEV感知网络模型泛化训练, 获得鲁棒的网络权重, 将训练得到的网络权重提供给网络模型进行推理使用。

在一些实施例中, 车载各视角周视针孔摄像机和环视鱼眼摄像机2D视觉图像, 实现传感器时间戳数据对齐配对时, 如曝光频率相同, 可考虑采用硬件同步触发对齐, 但若曝光不同频, 也可采用最近邻时间配对的方式, 一般视觉传感器具有较高的采样频率, 所以最近邻配对引入的时延误差影响一般可控。

在一些实施例中, 模型训练的总损失函数 L 由所有感知任务预测损失共同加权构成, 其中3D目标检测损失 L_{det} 一般可由分类Focal损失与3D框回归L1损失加权计算获得, 道路可通行区域分割损失 L_{seg} 一般可由交叉熵计算获得, 而车道线路沿检测损失则可由L1损失计算关键点及几何中心位置偏差获得, 网络训练用于提升泛化性的随机数据增强方法包括但不限于随机尺度变换、区域裁切、对称镜像、色彩变换、网格蒙板等。

基于前述的实施例, 本公开实施例提供一种车辆的环境检测装置, 如图12所示, 车辆的环境检测装置1200包括: 特征提取部1201, 配置为分别对车辆的周视图像和环视图像进行特征提取, 得到周视图像特征和环视图像特征; 第一确定部1202, 配置为基于检测任务对应的任务类型, 确定第一查询信息; 第二确定部1203, 配置为基于所述周视图像特征和所述环视图像特征, 确定第一键信息和第一值信息; 处理部1204, 配置为利用注意力机制, 对所述第一查询信息、所述第一键信息和所述第一值信息进行处理, 得到目标融合特征; 第三确定部1205, 配置为基于所述目标融合特征, 对所述车辆的周围环境进行检测, 确定与所述检测任务对应的检测结果。

在一些实施例中,所述周视图像特征包括二维周视特征和三维周视位置特征,所述环视图像特征包括二维环视特征和三维环视位置特征;特征提取部包括:第一编码器件,配置为对所述周视图像和所述环视图像进行二维编码,确定所述二维周视特征和所述二维环视特征;第二编码器件,配置为对所述周视图像和所述环视图像进行三维编码,确定所述三维周视位置特征和所述三维环视位置特征。

在一些实施例中,第一确定部包括:第一确定器件,配置为基于所述检测任务对应的任务类型和各历史时刻所述检测任务对应的检测结果,确定所述检测任务对应的第二查询信息、第二键信息和第二值信息;处理器件,配置为利用所述注意力机制,对所述第二查询信息、所述第二键信息和所述第二值信息进行处理,得到所述第一查询信息。

在一些实施例中,第一确定器件包括:第一确定子器件,配置为基于各所述历史时刻与当前时刻所述车辆的位姿转换关系,对各所述历史时刻所述检测任务对应的检测结果进行重构,确定各所述历史时刻所述检测任务对应的重构特征;第二确定子器件,配置为基于所述检测任务对应的任务类型和第一历史时刻所述检测任务对应的重构特征,确定当前时刻所述检测任务对应的初始重构特征;所述第一历史时刻为当前时刻的上一时刻;第三确定子器件,配置为将所述当前时刻所述检测任务对应的初始重构特征确定为所述第二查询信息;第四确定子器件,配置为基于各第二历史时刻所述检测任务对应的重构特征和所述当前时刻所述检测任务对应的初始重构特征,确定所述第二键信息和所述第二值信息;所述第二历史时刻为除所述第一历史时刻之外的历史时刻。

在一些实施例中,第一确定子器件还配置为:基于所述位姿转换关系,将各所述历史时刻所述检测任务对应的任务信息对齐至所述当前时刻,确定各所述历史时刻所述检测任务对应的更新检测结果;基于所述位姿转换关系、各所述历史时刻与所述当前时刻之间的时间差、所述检测任务对应的运动速度,对各所述历史时刻所述检测任务对应的更新检测结果进行重构,确定各所述历史时刻所述检测任务对应的重构特征。

在一些实施例中,第二确定子器件,还配置为:基于所述第一历史时刻所述检测任务对应的重构特征,确定所述当前时刻所述检测任务对应的第一初始重构特征;基于所述检测任务对应的任务类型,随机生成所述当前时刻所述检测任务对应的第二初始重构特征;将所述第一初始重构特征和所述第二初始重构特征进行拼接,确定所述当前时刻对应的初始重构特征。

在一些实施例中,第二确定子器件,还配置为:基于所述检测任务对应的任务类型,随机生成当前时刻所述检测任务对应的初始特征;对所述检测任务对应的初始特征进行重构,确定当前时刻所述检测任务对应的第二初始重构特征。

在一些实施例中,第二确定子器件,还配置为以下至少之一:在所述检测任务中包括运动目标检测任务的情况下,对三维空间的坐标进行均匀采样,确定所述运动目标检测任务对应的初始特征;在所述检测任务中包括道路分割检测任务的情况下,基于设定的车辆可通行区域中各子区域的中心位置的坐标,确定所述道路分割检测

任务对应的初始特征；在所述检测任务中包括车道线检测任务的情况下，对二维空间的坐标进行均匀采样，确定所述车道线检测任务对应的初始特征。

在一些实施例中，第二确定部包括：第二确定器件，配置为将所述二维周视特征和所述二维环视特征进行拼接，确定所述检测任务对应的第一键信息；拼接器件，配置为将所述二维周视特征、所述三维周视位置特征、所述二维环视特征和所述三维环视位置特征进行拼接，确定所述检测任务对应的第一值信息。

在一些实施例中，处理部包括：生成器件，配置为将所述第一查询信息作为所述注意力机制的查询矩阵、所述第一键信息作为所述注意力机制的键矩阵，生成注意力矩阵；第三确定器件，配置为基于所述注意力矩阵和所述第一值信息，确定所述目标融合特征。

在一些实施例中，第一确定部包括：第四确定器件，配置为针对每一所述检测任务，基于所述检测任务对应的任务类型，确定所述检测任务对应的查询信息；第五确定器件，配置为将各所述检测任务对应的查询信息进行拼接，确定所述第一查询信息。

以上装置实施例的描述，与上述方法实施例的描述是类似的，具有同方法实施例相似的有益效果。在一些实施例中，本公开实施例提供的装置具有的功能或包含的结构可以用于执行上述方法实施例描述的方法，对于本公开实施例的装置实施例中未披露的技术细节，请参照本公开实施例的方法实施例的描述而理解。

本公开实施例还提出了一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机指令，所述处理器执行所述指令时实现上述方法中的部分或全部步骤。

本公开实施例还提出了一种计算机可读存储介质，其上存储有计算机指令，所述计算机指令被处理器执行时实现上述方法中的部分或全部步骤。所述计算机可读存储介质可以是瞬时性的，也可以是非瞬时性的。

本公开实施例还提出了一种计算机程序，包括计算机可读代码，在所述计算机可读代码在计算机设备中运行的情况下，所述计算机设备中的处理器执行用于实现上述方法中的部分或全部步骤。

本公开实施例还提出了一种计算机程序产品，所述计算机程序产品包括存储了计算机指令的非瞬时性计算机可读存储介质，所述计算机指令被计算机读取并执行时，实现上述方法中的部分或全部步骤。该计算机程序产品可以通过硬件、软件或其结合的方式实现。在一些实施例中，所述计算机程序产品体现为计算机存储介质，在另一些实施例中，计算机程序产品体现为软件产品，例如软件开发包（Software Development Kit, SDK）等等。

上文对各个实施例的描述倾向于强调各个实施例之间的不同之处，其相同或相似之处可以互相参考。以上设备、存储介质、计算机程序及计算机程序产品实施例的描述，与上述方法实施例的描述是类似的，具有同方法实施例相似的有益效果。对于本公开实施例的设备、存储介质、计算机程序及计算机程序产品实施例中未披露的技术细节，请参照本公开实施例的方法实施例的描述而理解。

本公开实施例提供一种计算机设备的硬件实体，如图13所示，计算机设备1300的硬件实体包括：处理器1301

通常控制计算机设备1300的总体操作。通信接口1302可以使计算机设备通过网络与其他终端或服务器通信。存储器1303配置为存储由处理器1301可执行的指令和应用,还可以缓存待处理器1301以及计算机设备1300中各部待处理或已经处理的数据(例如,图像数据、音频数据、语音通信数据和视频通信数据),可以通过闪存(FLASH)或随机访问存储器(Random Access Memory, RAM)实现。处理器1301、通信接口1302和存储器1303之间可以通过总线1304进行数据传输。

可选地,说明书通篇中提到的“一个实施例”或“一实施例”意味着与实施例有关的特定特征、结构或特性包括在本公开实施例的至少一个实施例中。因此,在整个说明书各处出现的“在一个实施例中”或“在一实施例中”未必一定指相同的实施例。此外,这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。可选地,在本公开实施例的各种实施例中,上述各步骤/过程的序号的大小并不意味着执行顺序的先后,各步骤/过程的执行顺序可基于其功能和内在逻辑确定,而不对本公开实施例的实施过程构成任何限定。上述本公开实施例序号仅为描述,不代表实施例的优劣。

在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

在本公开实施例中,所揭露的设备和方法,可以通过其它的方式实现。以上所描述的设备实施例仅是示意性的,例如,单元的划分,仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,如:多个单元或组件可以结合,或可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以是通过一些接口,设备或单元的间接耦合或通信连接,可以是电性的、机械的或其它形式的。

上述作为分离部件说明的单元可以是、或也可以不是物理上分开的,作为单元显示的部件可以是、或也可以不是物理单元;既可以位于一个地方,也可以分布到多个网络单元上;可以根据实际的需要选择其中的部分或全部单元来实现本公开实施例方案。另外,在本公开实施例中的各功能单元可以全部集成在一个处理单元中,也可以是各单元分别单独作为一个单元,也可以两个或两个以上单元集成在一个单元中;上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。

本领域普通技术人员可以理解:实现上述方法实施例的全部或部分步骤可以通过指令相关的硬件来完成,前述的指令可以存储于计算机可读取存储介质中,该指令在执行时,执行包括上述方法实施例的步骤;而前述的存储介质包括:移动存储设备、只读存储器(Read Only Memory, ROM)、磁碟或者光盘等各种可以存储程序代码的介质。

或者，本公开实施例的上述集成的单元以软件功能模块的形式实现并作为独立的产品销售或使用，也可以存储在一个计算机可读取存储介质中。基于这样的理解，本公开实施例的技术方案本质上或者说做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机、服务器、或者网络设备）执行本公开实施例方法的全部或部分。而前述的存储介质包括：移动存储设备、ROM、磁碟或者光盘等各种可以存储程序代码的介质。

以上实施例仅是为充分说明本公开实施方式，本公开实施例的保护范围不限于此。本技术领域的技术人员在本公开实施例基础上所作的等同替代或变换，均在本公开实施例的保护范围之内。

权利要求书

1、一种车辆的环境检测方法，包括：

分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征；

基于检测任务对应的任务类型，确定第一查询信息；

基于所述周视图像特征和所述环视图像特征，确定第一键信息和第一值信息；

利用注意力机制，对所述第一查询信息、所述第一键信息和所述第一值信息进行处理，得到目标融合特征；

及

基于所述目标融合特征，对所述车辆的周围环境进行检测，确定与所述检测任务对应的检测结果。

2、根据权利要求1所述的方法，其中，所述周视图像特征包括二维周视特征和三维周视位置特征，所述环视图像特征包括二维环视特征和三维环视位置特征；

所述分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征，包括：

对所述周视图像和所述环视图像进行二维编码，确定所述二维周视特征和所述二维环视特征；及

对所述周视图像和所述环视图像进行三维编码，确定所述三维周视位置特征和所述三维环视位置特征。

3、根据权利要求1所述的方法，其中，所述基于检测任务对应的任务类型，确定第一查询信息，包括：

基于所述检测任务对应的任务类型和各历史时刻所述检测任务对应的检测结果，确定所述检测任务对应的第二查询信息、第二键信息和第二值信息；及

利用所述注意力机制，对所述第二查询信息、所述第二键信息和所述第二值信息进行处理，得到所述第一查询信息。

4、根据权利要求3所述的方法，其中，所述基于所述检测任务对应的任务类型和各历史时刻所述检测任务对应的检测结果，确定所述检测任务对应的第二查询信息、第二键信息和第二值信息，包括：

基于各所述历史时刻与当前时刻所述车辆的位姿转换关系，对各所述历史时刻所述检测任务对应的所述检测结果进行重构，确定各所述历史时刻所述检测任务对应的重构特征；

基于所述检测任务对应的任务类型和第一历史时刻所述检测任务对应的重构特征，确定所述当前时刻所述检测任务对应的初始重构特征；其中，所述第一历史时刻为所述当前时刻的上一时刻；

将所述当前时刻所述检测任务对应的所述初始重构特征确定为所述第二查询信息；及

基于各第二历史时刻所述检测任务对应的重构特征和所述当前时刻所述检测任务对应的所述初始重构特征，确定所述第二键信息和所述第二值信息；所述第二历史时刻为除所述第一历史时刻之外的历史时刻。

5、根据权利要求4所述的方法，其中，所述基于各所述历史时刻与当前时刻所述车辆的位姿转换关系，对各所述历史时刻所述检测任务对应的检测结果进行重构，确定各所述历史时刻所述检测任务对应的重构特征，包括：

基于所述位姿转换关系,将各所述历史时刻所述检测任务对应的检测结果对齐至所述当前时刻,确定各所述历史时刻所述检测任务对应的更新检测结果;及

基于所述位姿转换关系、各所述历史时刻与所述当前时刻之间的时间差、所述检测任务对应的运动速度,对各所述历史时刻所述检测任务对应的所述更新检测结果进行重构,确定各所述历史时刻所述检测任务对应的所述重构特征。

6、根据权利要求4所述的方法,其中,所述基于所述检测任务对应的任务类型和第一历史时刻所述检测任务对应的重构特征,确定当前时刻所述检测任务对应的初始重构特征,包括:

基于所述第一历史时刻所述检测任务对应的所述重构特征,确定所述当前时刻所述检测任务对应的第一初始重构特征;

基于所述检测任务对应的任务类型,随机生成所述当前时刻所述检测任务对应的第二初始重构特征;及

将所述第一初始重构特征和所述第二初始重构特征进行拼接,确定所述当前时刻对应的所述初始重构特征。

7、根据权利要求6所述的方法,其中,所述基于所述检测任务对应的任务类型,随机生成所述当前时刻所述检测任务对应的第二初始重构特征,包括:

基于所述检测任务对应的所述任务类型,随机生成所述当前时刻所述检测任务对应的初始特征;及

对所述检测任务对应的所述初始特征进行重构,确定所述当前时刻所述检测任务对应的所述第二初始重构特征。

8、根据权利要求7所述的方法,其中,所述基于所述检测任务对应的任务类型,随机生成当前时刻所述检测任务对应的初始特征,包括以下至少之一:

在所述检测任务中包括运动目标检测任务的情况下,对三维空间的坐标进行均匀采样,确定所述运动目标检测任务对应的所述初始特征;

在所述检测任务中包括道路分割检测任务的情况下,基于设定的车辆可通行区域中各子区域的中心位置的坐标,确定所述道路分割检测任务对应的所述初始特征;及

在所述检测任务中包括车道线检测任务的情况下,对二维空间的坐标进行均匀采样,确定所述车道线检测任务对应的所述初始特征。

9、根据权利要求2所述的方法,其中,所述基于所述周视图像特征和所述环视图像特征,确定第一键信息和第一值信息,包括:

将所述二维周视特征和所述二维环视特征进行拼接,确定所述检测任务对应的所述第一键信息;及

将所述二维周视特征、所述三维周视位置特征、所述二维环视特征和所述三维环视位置特征进行拼接,确定所述检测任务对应的所述第一值信息。

10、根据权利要求1至9任一项所述的方法,其中,所述利用注意力机制,对所述第一查询信息、所述第一

键信息和所述第一值信息进行处理，得到目标融合特征，包括：

将所述第一查询信息作为所述注意力机制的查询矩阵、所述第一键信息作为所述注意力机制的键矩阵，生成注意力矩阵；及

基于所述注意力矩阵和所述第一值信息，确定所述目标融合特征。

11、根据权利要求1至9任一项所述的方法，其中，所述检测任务的数量为多个；所述基于检测任务对应的任务类型，确定第一查询信息，包括：

针对每一所述检测任务，基于所述检测任务对应的任务类型，确定所述检测任务对应的查询信息；及

将各所述检测任务对应的查询信息进行拼接，确定所述第一查询信息。

12、根据权利要求1至9任一项所述的方法，还包括：

对所述车辆的雷达点云数据进行特征提取，得到点云特征；

所述基于所述周视图像特征和所述环视图像特征，确定第一键信息和第一值信息，包括：

基于所述周视图像特征、所述环视图像特征和所述点云特征，确定所述第一键信息和所述第一值信息。

13、一种车辆的环境检测装置，包括：

特征提取部，配置为分别对车辆的周视图像和环视图像进行特征提取，得到周视图像特征和环视图像特征；

第一确定部，配置为基于检测任务对应的任务类型，确定第一查询信息；

第二确定部，配置为基于所述周视图像特征和所述环视图像特征，确定第一键信息和第一值信息；

处理部，配置为利用注意力机制，对所述第一查询信息、所述第一键信息和所述第一值信息进行处理，得到目标融合特征；及

第三确定部，配置为基于所述目标融合特征，对所述车辆的周围环境进行检测，确定与所述检测任务对应的检测结果。

14、一种计算机设备，包括存储器和处理器，所述存储器存储有可在处理器上运行的计算机指令，所述处理器执行所述指令时实现权利要求1至12任一项所述的方法。

15、一种车辆，包括如权利要求14所述的计算机设备。

16、一种计算机可读存储介质，其上存储有计算机指令，该计算机指令被处理器执行时实现权利要求1至12任一项所述的方法。

17、一种计算机程序产品，包括计算机指令，该计算机指令被处理器执行时，实现权利要求1至12中任一项所述的方法。